IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

# Fast Force-Closure Grasp Synthesis with Learning-Based Sampling

Wei Xu, Weichao Guo, Xu Shi, Xinjun Sheng*, and Xiangyang Zhu

***Abstract*—Anthropomorphic robotic hands have been widely investigated to dexterously manipulate objects because of their anatomical similarity to the human hand. However, the large dimension of configuration space challenges the real-time performance of existing grasp planning methods and drastically limits the application of anthropomorphic hands. In this letter, we propose a fast force-closure grasp synthesis (FFCGS) method for the anthropomorphic hand to efficiently grasp unknown objects. The FFCGS is implemented by using a signed distance field (SDF) as input. Firstly, a network that samples feasible 6D wrist poses is trained in an end-to-end fashion to reduce the dimension of search space. Furthermore, a fast optimization algorithm is presented to find finger configurations for force-closure precision grasp based on the differentiable Q-distance metric. We validate our method in both a simulated and a real-world environment. Experiment results show that the proposed FFCGS achieves a significantly improved performance in terms of time efficiency (5 times faster), grasp quality metrics, and success rate (5%-10% improvement) over benchmark methods. The outcomes of this study have great significance in promoting the motion planning of robot hand-arm systems and upper-limb prostheses.**

***Index Terms*—Multifingered hands, grasping, deep learning in grasping and manipulation, force-closure, learning-based sampling.**

## I. Introduction

ANTHROPOMORPHIC hands with more actuated joints have higher dexterity than parallel grippers. This gives anthropomorphic hands potential in situations where advanced manipulability is required, such as service robots and prosthetic hands [1]. Researchers have explored many methods to reduce the degree of actuation, such as novel underactuated mechanisms [2]. However, grasp planning by optimization still faces considerable computation complexity caused by the high dimension of the multi-fingered hand. Furthermore, a smooth collaboration of robotic arms and hands usually requires multiple grasps with diverse orientations as fast as possible [3].

Previous works have proposed different approaches to speed up grasp synthesis. The force-closure optimization problem is approximated into a format suitable for fast computation in [4]. [5] simplifies the optimization problem by relaxing the friction cone constraint. Some try to divide the whole optimization into problems with fewer dimensions and constraints. In [6], the object point cloud is preprocessed into a hierarchical structure that allows multi-stage optimization of hand configurations. Similarly, [7] optimizes the palm pose and joint angles in two separate optimization algorithms to improve the convergence properties. Researchers also explore object geometry and demonstrations to narrow the scope of grasp synthesis. In [8], [9], the contact points detected by the neural network model are used as constraints for the subsequent grasp synthesis algorithm. [10] approximate objects into primitive shapes before using heuristic rules to generate grasps candidates, which are further refined by planners. Despite the effort, planning both the wrist pose and joint angles for anthropomorphic hands still costs from several seconds to a minute, which is unsuitable for real-time applications.

In this letter, we propose a method to plan precision grasp for anthropomorphic hands that draw inspiration from previous works. The grasp synthesis problem is decomposed into two separate parts that can be solved efficiently. We start with a lightweight generative model that learns to extract features from object signed distance field (SDF) and predict the translation and rotation parameters of the wrist. A dataset with 50k grasps on 50 objects is annotated in a simulator to train the network. We can significantly reduce the search space dimension and the number of possible local optimums of the following finger joint optimization by fixing the wrist pose. Then, a fast and differentiable force-closure metric named Q-distance is used to find the optimal finger joint angles that can grasp the object stably [11]. In addition, the SDF representation with additional geometry information than the point cloud is believed to improve the performance of wrist pose sampling [12]. It allows efficient grasp optimization without the need to compute expensive surface normals and distances using point clouds. We evaluated our approach in both simulation and actual world experiments. Results demonstrate that our method can simultaneously output dozens of stable grasps with diverse directions in a second.

In summary, our primary contributions are: (1) A novel generative network that fully utilizes the object features to sample feasible wrist poses. (2) A fast optimization algorithm based on the differentiable Q-distance metric that searches for the optimal force-closure grasps. (3) We use a fast two-stage grasp synthesis framework that simplifies the high-dimensional optimization problem by using learning-based wrist pose sampling to guide finger joint optimization.

Manuscript received: March, 27, 2023; Accepted May, 23, 2023.

This paper was recommended for publication by Editor Hong Liu upon evaluation of the Associate Editor and Reviewers' comments.

This work is supported in part by the the National Natural Science Foundation of China (Grant Nos. 91948302, 52175021), in part by Natural Science Foundation of Shanghai (Grant No. 23ZR1429700).

The authors are with the State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China and Meta Robotics Institute, Shanghai Jiao Tong University, Shanghai 200240, China. (e-mail: xu.wei@sjtu.edu.cn; guoweichao90@gmail.com; sjtu-shixu@sjtu.edu.cn; xjsheng@sjtu.edu.cn; mexyzhu@sjtu.edu.cn)

Digital Object Identifier (DOI): see top of this page.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.**

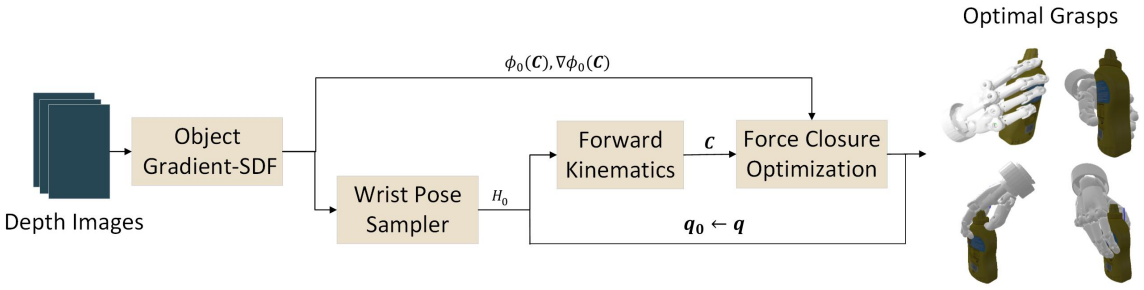


Fig. 1. Planning dexterous grasps for anthropomorphic hands. Depth images are integrated into the SDF and pass through the wrist pose sampler. Then, the force-closure optimization takes the predicted pose $H_0$, contacts $\mathbf{C}$, distance $\phi_O(\mathbf{C})$ and normals $\nabla\phi_O(\mathbf{C})$ of the object as input to iteratively update the joint configurations that produce force-closure grasps.

## II. Related Work

### A. Grasp Optimization

Previous works have achieved significant progress in grasping known objects using an analytical approach that explicitly models the interaction between the object and the hand, then finds optimal hand configurations using quality metrics [13]. Most grasp metrics quantify the property of force closure, which means a grasp can immobilize the object by balancing any external force, such as the $\epsilon$ metric [14].

While optimization algorithms using these metrics can generate force-closure grasps, the efficiency can be further improved. [5] proposes a quick grasp synthesis algorithm that simplifies the force-closure criteria by ignoring contact friction. [15] formulates a generalized force-closure metric that penalizes the distance between the fingertip and the object surface to relax the contact constraint. [16] uses a fast support function to compute the largest-minimum resistible wrench. In [17], a dual-stage iterative optimization strategy is used to search separately for palm and finger configuration. [7] presents a power grasp planning method that first proposes grasp candidates based on heuristic rules using local geometry features. Then the shape complementary metric that penalizes the distance between the gripper points and the object points is used to refine grasps further. The computational complexity caused by the high dimension still limits the application of grasp optimization methods.

Our method has two key differences compared to existing methods. (1) Differentiable optimization objectives that explicitly consider friction. It makes grasp optimization faster by simply solving linear programming problems for gradients of the objectives and applying gradient descent rules. The grasp optimization can converge within dozens of iterations. (2) An efficient optimization algorithm that starts with initial guesses by sampling on the learned posterior distribution of successful grasps. We don't use computationally expensive population-based optimization algorithms (swarm, particle, stochastic algorithms), which are common in existing grasp optimization methods to prevent stuck in local minima because of the multimodality of the grasp objective function [18]. It speeds up the grasp synthesis time of optimization-based methods from seconds to dozens of milliseconds while not sacrificing the quality of grasps.

### B. Learning-based Grasp synthesis

In recent years, learning-based methods have made remarkable progress in the robotic grasping field. These methods train a network that learns from successful grasp datasets collected in actual or simulated environments, and estimates grasp directly from raw sensor input [19]. However, most works are limited to parallel grippers because of the simplicity of generating training data and small DoFs [20].

Few works are focused on the multi-fingered hand. [21] first trains a network to take RGBD images as input to predict contact points and uses a grasp planner to retrieve the finger joints of the multi-fingered hand. In [8], a multi-stage model that selects one contact point at each stage is followed by an inverse kinematics solver. Unlike previous viewpoint-dependent methods, [22] presents a generative model that samples hand configurations given RGBD images and a fixed grasp orientation. In [23], a generative model is trained with 1.5M annotated grasps, which consists of a variational grasp generator and an iterative refine module. It takes an object point cloud as input and samples possible grasps before iterative refinement for anthropomorphic hands. End-to-end learning-based multi-fingered planning methods usually require a massive dataset with millions of grasps that are time-consuming to collect in simulators. Moreover, most existing approaches need a iterative refinement module that only penalizes the contact loss between the hand and the object without considering grasp stability. It makes them unsuitable to produce precision grasps, which are essential for subsequential dexterous manipulations [24].

In [25], the authors took a unique approach by formulating learned grasp success probability and conditional prior models as objectives, then planned grasps using a non-linear optimization solver. Similarly, we use a network to model the distribution of good grasps, then sample the distribution as initial guesses for efficient optimization. However, our approach has two key differences. First, our optimization objective is based on analytical modeling, which ensures generalization ability across objects and the force-closure property. Second, our optimization is based on solving linear programming problems instead of non-linear problems, which are usually much easier to solve.

**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.**

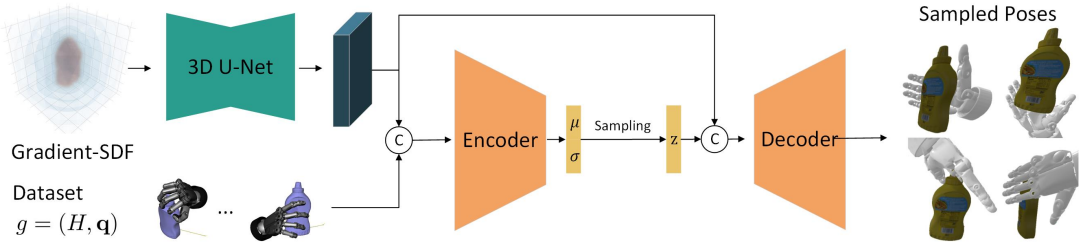


Fig. 2. The network structure of the wrist pose sampler. The proposed model is trained on a grasp dataset annotated by a simulator to take the Gradient-SDF as input and generate a diverse set of wrist poses.

## III. Problem Formulation

The input to our method is a signed distance field (SDF) $\phi_O(x)$ of the object $\mathcal{O}$, a discrete voxel grid that stores the signed closest distance from $x$ to the object. SDF can be reconstructed in real-time from a single depth sensor [26], and further refined by applying a 3D completion network [27]. A grasp of the anthropomorphic hand is defined as

$$g = (H, \mathbf{q}) | H \in SE(3), \mathbf{q} \in \mathbb{R}^d \tag{1}$$

where $\mathbf{q} = [q_1, ..., q_d]$ is the vector of hand configurations, $d$ is the number of DoF, $H$ is the wrist pose with rotation $R \in SO(3)$ and translation $\mathbf{t} \in \mathbb{R}^3$.

The pipeline of our method that generates dexterous grasps for multi-fingered hands is shown in Fig.1. It consists of two submodules: Wrist Pose Sampler and Gradient-Based Grasp Optimization.

## IV. Method

### A. Wrist Pose Sampler

The task of the wrist pose sampler is to sample from a posterior distribution $P(H^*|\Phi)$, where $H^*$ is the wrist pose that leads to successful grasps, and $\Phi$ represents the object SDF. For this purpose, we propose a generative model, as shown in Fig.2.

We use the Gradient-SDF as input, a hybrid representation of the object that stores both the signed distance and its gradient vector in each voxel [28]. This representation is a four-channel voxel grid that contains more local geometry information, which helps to reduce the reconstruction cost by providing more local information compared to the simple SDF representation. A modified 3D U-Net with residual blocks is used to extract local and global features without the vanishing gradient problem. The size of the output features is $32 \times 16^3$.

At the training stage, the grasp configuration $g$ is concatenated to the feature vector at each voxel to become the input to Fully Convolutional Network (FCN) based encoder which further help to reduce the orientational reconstruction cost. The encoder predicts the distribution $\mathcal{N}(\mu, \sigma)$ of the two-dimensional latent variable $z$. Then, the decoder takes the concatenation of latent variable $z$ and the features as input to reconstruct the grasp configurations $\hat{g}$. In actual use, the encoder is discarded. The decoder uses the features and the latent variable $z$ sampled from $\mathcal{N}(0, I)$ to predict wrist poses.

The encoder and the decoder consist of 4 convolutional layers interleaved with pooling layers with a stride of 2, followed by different heads of Multi-layer Perceptron (MLP) to predict the distribution, the translation, and the rotation. The entire network is trained end-to-end on ground-truth wrist poses obtained in simulation. The total loss function is defined as follows:

$$\mathcal{L}_{tot} = \mathcal{L}_R + \beta_t \mathcal{L}_{KL} \tag{2}$$

$\mathcal{L}_R$ represents the reconstruction loss of wrist poses. We simplify the computation of reconstruction loss by only considering the average Euclidean distances between the predefined vertices of the original $p = p_1, ..., p_m$ and reconstructed $\hat{p} = \hat{p}_1, ..., \hat{p}_m$ tetrahedrons fixed on the palm, instead of the entire hand meshes.

$$\mathcal{L}_R = \frac{1}{m} \sum_{i}^{m} ||p_i - \hat{p}_i||_2 \tag{3}$$

$\mathcal{L}_{KL}$ is the Kullback–Leibler divergence loss that penalizes the difference between the latent distribution and a standard Gaussian Distribution. To avoid the KL vanishing problem, we adopt the cyclical annealing schedule proposed in [29], in which the coefficient $\beta_t$ changes from 0 to 0.005 cyclically.

$$\mathcal{L}_{KL} = -KL(p(z|\phi, H)||\mathcal{N}(0, I)) \tag{4}$$

We select 50 objects from the YCB object set [30], including all the food, kitchen, and tool items, 20% were held out as a testing set, and 80% as a training set. In addition, the EGAD! Eval set of 49 diverse objects is used in later experiments to validate the generalization ability of our model [31]. The Gradient-SDF is generated by differentiating the SDF volume computed from the object mesh using Open3D [32]. We use a simulated annealing grasp planner in GraspIt [33]. Two hundred grasps with the best $\epsilon$ value are annotated for each object. Then, we add zero-mean Gaussian noise to the translation and rotation parameters of the annotated grasps. In addition, we apply five random rotations to each grasp and its corresponding Gradient-SDF to get the final dataset of 50k grasps. A mainstream desktop computer is used to train and validate the network, which consists of an Intel i7 CPU and an NVIDIA GeForce RTX 2060 GPU. We use the Adam optimizer to train the network: the learning rate is $10^{-4}$, five-fold cross-validation is applied, the batch size is set to 32, and the epoch is set to 100.

**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.**

### B. Gradient-based Grasp Optimization

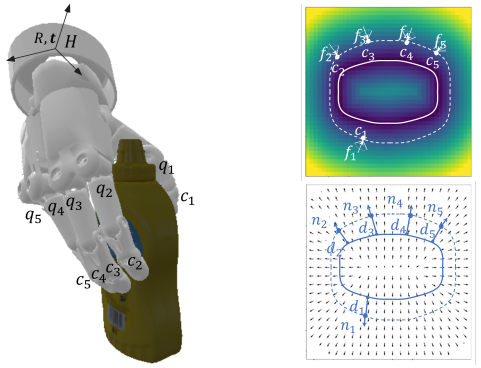


Fig. 3. Illustration of the grasping problem. (a) The kinematic parameters of the anthropomorphic hand. (b) Cross section of the object. The line solid ellipses are the surface of the object. The dotted line ellipse passes through the fingertip contact points $\mathbf{c}_i$ and their corresponding grasp forces $\mathbf{f}_i$, normals $\mathbf{n}_i$ and distances $d_i$.

The grasp optimization problem is illustrated in Fig.3. A set of wrist poses $\{H\}$ are predicted by the wrist sampler and are filtered to avoid collisions. Then the grasp optimizer searches for the joint configurations $\mathbf{q}$ that form stable grasps of the object. The optimization problem is given in (5). $E$ is a differentiable quality function, where $E_g$ quantifies the grasp quality, $E_c$ penalizes improper contacts, and $E_q$ is used to avoid unnatural joint configurations.

$$\begin{aligned} \min_{\mathbf{q}} \quad & E = E_g(\mathbf{C},\mathbf{N}) + a_c E_c(\mathbf{C},\mathbf{d}) + a_q E_q(\mathbf{q}) \\ s.t. \quad & \mathbf{q} \in [\mathbf{q}_{min}, \mathbf{q}_{max}] \end{aligned} \tag{5}$$

where $\mathbf{C} = [\mathbf{c}_1, ..., \mathbf{c}_n]$ is a set of contacts on the object $\mathcal{O}$, the contacts are punctual contacts with friction [34]; $\mathbf{N} = [\mathbf{n}_1, ..., \mathbf{n}_n]$ are the surface normals at the contacts, $\mathbf{d} = [d_1, ..., d_n]$ are the distance between fingertips and surface, and $a_c, a_q$ are set to 5 and 0.5 to penalize large fingertip distance to the object surface and joint configuration approaching limits. The possible wrench that robotic fingers can produce at each contact point can be defined as

$$w(\mathbf{f}_i, \mathbf{c}_i) = [\mathbf{f}_i, \mathbf{c}_i \times \mathbf{f}_i]^T, \quad \mathbf{f}_i \in FC(\mathbf{c}_i, \mathbf{n}_i, \mu) \tag{6}$$

where $\mathbf{n}_i$ is the outward-pointing normal vector of the object surface at contact $\mathbf{c}_i$, $\mathbf{f}_i$ is the force applied at the contact, and the coefficient of friction $\mu$ is set to 0.2. We can linearize the friction cone $FC(\mathbf{c}_i, \mathbf{n}_i, \mu)$ using a $m$-sided polyhedral cone:

$$\hat{FC}(\mathbf{c}_i, \mathbf{n}_i) = \{\mathbf{f}_i | \mathbf{f}_i = \sum_j^m \alpha_{ij} T(\mathbf{c}_i, \mathbf{n}_i)\mathbf{e}_j\} \tag{7}$$

where $\mathbf{e}_j$ is a unit vector uniformly distributed on the standard friction cone that has $z$-axis as its axis and the origin as its origin, $T(\mathbf{c}_i, \mathbf{n}_i)$ is the transformation that aligns the standard frictional cone to the current contact $\mathbf{c}_i$ and its normal $\mathbf{n}_i$, $\alpha_{ij}$ are non-negative coefficients. The possible overall wrench that the robotic hand can produce to compensate for the external load without slippery can be represented by a convex hull known as Grasp Wrench Space (GWS) [35]

$$C_{\mathbf{w}} = \{\mathbf{w} | \mathbf{w} = \sum_i^n \sum_j^m \alpha_{ij}\mathbf{w}_{ij}, \sum_i^n \sum_j^m \alpha_{ij} \leq 1\} \tag{8}$$

where $\sum_i^n \sum_j^m \alpha_{ij} \leq 1$ to reflect the limits of finger forces, and the primitive wrench $\mathbf{w}_{ij} = w(T(\mathbf{c}_i, \mathbf{n}_i)\mathbf{e}_j, \mathbf{c}_i)$ is the wrench that $\mathbf{e}_j$ applies at the contact point $\mathbf{c}_i$. The GWS can be written as $C_{\mathbf{w}} = \{\mathbf{w} | \mathbf{w} = \sum_i^n \alpha_i \mathbf{w}_i, \sum_i^n \alpha_i \leq 1\}$. The force closure condition is achieved only when $0 \in C_{\mathbf{w}}$. The grasp stability can be defined as the distance between the origin and the $C_{\mathbf{w}}$. We linearize a unit sphere centered at the origin of the GWS $\hat{\mathcal{S}} = \{\sum_{k=1}^K \rho_k \mathbf{u}_k | \rho_k > 0\}$, where $\mathbf{u}_k$ is a unit vector uniformly distributed on the sphere, and $(\rho_1, \ldots, \rho_K)$ are the non-zero coordinates of the basis $(\mathbf{u}_1, \ldots, \mathbf{u}_K)$. The grasp quality Q-distance can be formulated as

$$\begin{aligned} & E_g = \min_{\mathbf{w}} \sum_{k=1}^K \rho_k \\ & s.t. \quad \sum_{k=1}^K \rho_k \mathbf{u}_k = \sum_i^n \alpha_i \mathbf{w}_i = \mathbf{w}, \mathbf{w} \in C_{\mathbf{w}} \end{aligned} \tag{9}$$

The $E_g$ measures the distance between the origin and the $C_{\mathbf{w}}$. When $E_g$ approaches zero, its corresponding grasp becomes closer to a force-closure grasp. The $E_g$ can be computed by linear programming method such as simplex algorithm, the resulting non-zero coefficients that satisfiy the constrain in (9) can be represented as $[\rho^*_{k_1}, ..., \rho^*_{k_r}, \alpha^*_{i_1}, ..., \alpha^*_{i_s}]$. The derivatives of $E_g$ can be obtained by differentiating both sides of the constrain in (9), as shown in (10).

$$\frac{\partial E_g}{\partial \mathbf{q}} = \sum_{k=1}^{k_r} \frac{\partial \rho^*_{k_i}}{\partial \mathbf{q}} = [\mathbf{1}_r, \mathbf{0}_{s-1}] D(\mathbf{q})^{-1} \sum_{k=1}^{s} \alpha^*_{i_k} \frac{\partial \mathbf{w}_{i_k}}{\partial \mathbf{q}} \tag{10}$$

$D(\mathbf{q}) = [\mathbf{u}_{k_1}, ..., \mathbf{u}_{k_r}, \mathbf{w}_{i_s} - \mathbf{w}_{i_1}, ..., \mathbf{w}_{i_s} - \mathbf{w}_{i_{s-1}}]$. Details can be found in our previous work [11]. Assume $\mathbf{w}_i$ is the wrench of the unit force $\mathbf{e}_j$ at contact $\mathbf{c}_i$. We simplify the local geometry to a plane at contacts, where $\frac{\partial \mathbf{n}_i(\mathbf{c}_i)}{\partial(\mathbf{c}_i)} = 0$. In addition, outward-pointing normal $\mathbf{n}_i$ at contact $\mathbf{c}_i$ can approximate by the gradient of object SDF $-\frac{\nabla\phi(\mathbf{c}_i)}{|\nabla\phi(\mathbf{c}_i)|}$. Then $\frac{\partial \mathbf{w}_i}{\partial \mathbf{q}}$ can be simplified as

$$\frac{\partial \mathbf{w}_i}{\partial \mathbf{q}} = \frac{\partial \mathbf{w}_i}{\partial \mathbf{c}_i} \frac{\partial \mathbf{c}_i}{\partial \mathbf{q}} = [\mathbf{0}, [\mathbf{e}_j]_\times]^T \mathbf{J}_{\mathbf{c}_i}(\mathbf{q}) \tag{11}$$

where $\mathbf{J}_{\mathbf{c}_i}(\mathbf{q})$ is the Jacobian of $\mathbf{c}_i$ from the Cartesian space to the joint space, such that $\mathbf{c}_i = FK(H, \mathbf{q})$. To relax the constraint $\mathbf{c}_i \in \partial\mathcal{O}$, we formulated contact quality $Q_c$ that penalizes penetration and distance between the fingertip and the object. The distance between the object and fingertips can be easily retrieved by querying the SDF function. Notice that a small penetration $d_0$ is usually helpful to grasp the object firmly and reduce the impact of sensor noise. The contact quality and its derivatives are given as

**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.**

$$
\begin{aligned}
E_c &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\phi(\mathbf{c}_i)+d_0)^2} \\
\frac{\partial E_c}{\partial \mathbf{q}} &= \sum_{i=1}^{n}\frac{\partial E_c}{\partial \mathbf{c}_i}\frac{\partial \mathbf{c}_i}{\partial \mathbf{q}} \\
&= \sum_{i=1}^{n}\frac{(\phi(\mathbf{c}_i)+d_0)\partial\phi(\mathbf{c}_i)/\partial \mathbf{c_i}}{\sqrt{n\sum_{i=1}^{n}(\phi(\mathbf{c}_i)+d_0)^2}}\mathbf{J}_{\mathbf{c}_i}(\mathbf{q})
\end{aligned}
\tag{12}
$$

where $d_0$ is set to 0.5cm in real-world experiments and 0 in simulated experiments, $\{\mathbf{c}_i\}$ is a set of predefined points on the fingertips. In addition, we add another penalty term to ensure the joint configurations stay within limits. The smaller $E_c$ is, the closer each contact is to the object's surface. The joint quality and its derivatives relative to $\mathbf{q}$ are defined as

$$
\begin{aligned}
E_q &= \frac{1}{d}\sum_{i=1}^{d}(\frac{q_i - q_{i0}}{q_{imax}-q_{imin}})^2 \\
\frac{\partial E_q}{\partial \mathbf{q}} &= \frac{2}{d}[\frac{q_1-q_{10}}{q_{1max}-q_{1min}},...,\frac{q_d-q_{d0}}{q_{dmax}-q_{dmin}}]
\end{aligned}
\tag{13}
$$

where $d$ is the degree of freedom, $q_{i0}, q_{imax}, q_{imin}$ are the mean, maximum, and minimum value of the joint $i$.

**Algorithm 1** Q-distance Grasp Optimization Algorithm.

**Input:** Object SDF $\phi$
**Initialization:** joint configurations $\mathbf{q} = \mathbf{0}$
  Sample wrist poses: $H_0 \leftarrow PoseFilter(Network(\phi))$.
  **for** $H \in H_0$ **do**
    **for** $t = 0 : T$ **do**
      Update contacts and their normals:
        $\mathbf{C}_t \leftarrow FK(H, \mathbf{q}), \mathbf{N}_t \leftarrow \nabla\phi(\mathbf{C}_t), \mathbf{d}_t \leftarrow \phi(\mathbf{C}_t)$.
      Solve for Q-distance and its coefficients:
        $E_g, [\rho^*_{k_1}, ..., \rho^*_{k_r}, \alpha^*_{i_1}, ..., \alpha^*_{i_s}] \leftarrow LP_{\rho,\alpha}(Q_g(\mathbf{C}_t, \mathbf{N}_t))$
      **if** $E_g < \epsilon$ **then**
        Break
      **end if**
      Update configurations: $\mathbf{q} \leftarrow \mathbf{q} - \beta_t \nabla_{\mathbf{q}} E(\mathbf{q}, \mathbf{C}_t, \mathbf{d}_t)$
    **end for**
    **if** not $ExistCollision(\mathbf{C}_0, ..., \mathbf{C}_T)$ **then**
      Break
    **end if**
  **end for**
  **Return:** $g = (H, \mathbf{q})$.

We use a gradient descent algorithm to optimize the joint configurations, as shown in Algorithm 1. A linear decay learning rate $\beta_t$ is used to mitigate noises near the object's surface. Collision-free and reachable grasps are filtered in two separate processes. Given predicted poses, we first filter out poses that will have vertexes of the hand's bounding box (joint angles are set to zeros) collide with the table and the object, which is achieved by querying the signed distance of each vertex. In addition, we discard poses with the palm pointing upwards since those poses are usually not reachable under our experimental setup. Then, we store the fingertip positions in the grasp optimization phase. Since the optimization starts from zero joint angles to optimal joint angles, the fingertip positions of each iteration step can be seen as an approximation of fingertip positions in actual grasping. We filter out grasps with cached fingertip positions lower than the table to prevent fingertips from colliding during grasping.

## V. EXPERIMENTS

Simulated and real-world experiments are executed to evaluate the proposed approach. The Schunk SVH hand, an anthropomorphic robot hand, is used in all the experiments.

### A. Simulated Experiments

*1) Performance of wrist pose sampler:* To evaluate the performance of our wrist pose sampler, we compare it with two baseline methods that are frequently used in the previous grasp synthesis literature: random sampling and geometric sampling [7], [19]. The YCB testing and EGAD eval set are merged and used in the simulation experiments.

To accommodate various sizes of the objects, we limit the translation $t$ to two times the size of the object's bounding box for both methods. The random method samples both translation and rotation parameters uniformly, similar to the initial condition for most population-based optimization algorithms [36]. The geometric sampling utilizes a heuristic idea that the normal of the palm usually aligns with the surface normal. This method uniformly samples the translation and uses the quaternions that align the palm normal with the gradient of SDF at the sampled translation; then, the sampled wrist poses are filtered for collision-free poses. Then all three methods sample 32 wrist poses for every object, followed by our grasp optimization algorithm up to 50 iterations to produce the final grasps. The total quality $E$, the grasp quality $E_g$, and the average sampling time are given in Table I.

TABLE I
RESULTS OF SAMPLING METHODS IN SIMULATION. ↓: LOWER THE BETTER

| Methods | Ours | Random Sampling | Geometric Sampling |
|---|---|---|---|
| $E\downarrow$ | **0.148** | 2.136 | 0.971 |
| $E_g\downarrow$ | **0.013** | 0.785 | 0.244 |
| Time (s)↓ | 3.68E-03 | **3.73E-06** | 5.00E-03 |

The results showed that our learning-based sampling method outperforms the other baseline methods in both two grasp metrics by a margin. The wrist poses sampled by our approach help to produce more stable grasps with a smaller distance to the GWS origin than the geometric sampling ($0.013$ vs. $0.244$), and more firm grasps with more accurate contacts ($0.148$ vs. $0.971$). The wrist pose sampler can learn from experiences to extract geometric features to predict likely wrist poses, which is more effective than the heuristic geometric and random sampling methods. Our method needs a shorter sampling time than the geometric sampling method because explicit collision checking is seldom needed using our volumetric network [12]. While random sampling uses the shortest time, our approach is the most cost-effective.

**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.**

*2) Performance of grasp optimization:* As shown in Fig.4, our grasp optimization algorithm can produce a reasonably good grasp in 10 iterations and converges to the optimal grasp with better contact quality after 50 iterations. The grasp optimization is evaluated under different resolutions of SDF. While a higher resolution is needed for a higher contact quality (less penetration or distance). It is shown that a higher density of grids has a negligible impact on the gesture of optimized grasps.

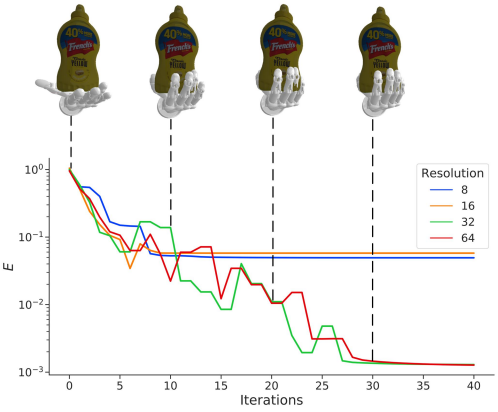


Fig. 4. Grasp optimization process with different SDF resolutions.

We compare the performance of our approach with GraspIt. Both methods generate 32 grasps for each object in the evaluation sets, from which 20 grasps with the highest grasp quality are selected to compute the average grasp qualities. To balance performance and speed, we set the resolution to 32 and the number of iterations to 50 in the following experiments. The histograms of the grasps obtained by both methods are shown in Fig. 5. We use the Wilcoxon signed-rank test to determine whether the methods differ statistically. Results show that our approach significantly improves grasp and contact quality over the benchmark ($p < 0.001$). In addition, we investigate the performance of both methods on each object set. The results are listed in Table II. Notice that the simulated annealing algorithm used in Graspit has relatively consistent performance in the YCB testing and EGAD eval sets in terms of $E$ and $E_c$. Results show that our approach performs better on both object sets than the Graspit except $E_g$ on the YCB testing set. It demonstrates that our approach can generalize to objects from different datasets with diverse complexity and difficulty.

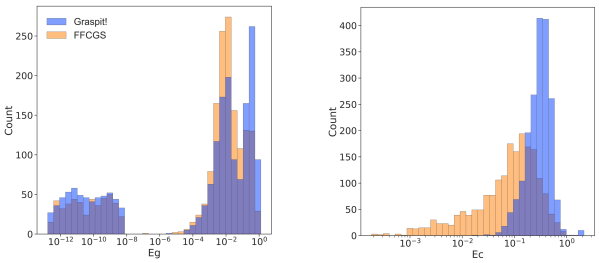


Fig. 5. The results produced by our method and the baseline in histograms. (a) The grasp quality $E_g$. (b) The contact quality $E_c$ (cm).

TABLE II
RESULTS OF GRASP SYNTHESIS METHODS. ↓: LOWER THE BETTER

| Object set | YCB testing | | EGAD! eval | |
|---|---|---|---|---|
| Planner | GraspIt! | Ours | GraspIt! | Ours |
| $E$ ↓ | 0.291 | **0.210** | 0.298 | **0.086** |
| $E_g$ ↓ | **0.014** | 0.017 | 0.026 | **0.009** |
| $E_c$ (cm) ↓ | 0.55 | **0.39** | 0.54 | **0.15** |

Fig.6 shows the grasps generated by the proposed method for the YCB and the EGAD eval objects. The results show that the sampler can predict reliable wrist poses from diverse directions, and our optimization algorithm can effectively find optimal force-closure grasps for various objects. However, our grasp synthesis can sometimes fail, as shown in Fig.6. The predicted wrist pose in Fig.7 (a) is about a centimeter away to grasp the scissor firmly. It is rare as the optimization can usually compensate for inaccurate wrist poses. In Fig.7 (b-c), the palm intersect with the object because we only penalize the penetration of fingertips for speed considerations, whose side effect can be reduced by using a reactive controller [37]. For tools with a small graspable area, interference between fingers may happen due to the mechanical design of the anthropomorphic hand, as shown in Fig.7 (d). It can be avoided by limiting the range of joint configurations.

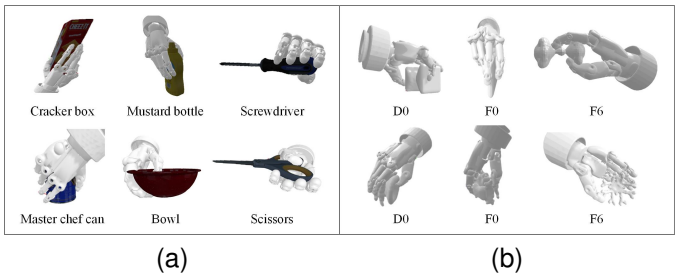


Fig. 6. Grasp examples of different objects with the multi-fingered hand. (a) YCB object set. (b) EGAD eval set.

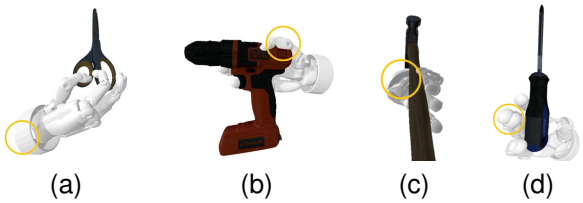


Fig. 7. Examples of failed grasps on tools. (a)Scissor. (b)Power drill. (c)Hammer. (d)Screwdriver.

## B. Real Robot Experiment

We set up a robot platform composed of a JAKA zu7 arm and a Schunk SVH hand to evaluate our approach in Fig.8. Images are captured from five angles using the RealSense D435i camera and the background is removed [39]. The SDF is computed by KinectFusion [40] before being used to generate grasps. The mid products of our method in grasping the Windex bottle are shown in Fig.9.

**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.**

TABLE III
REAL WORLD EXPERIMENTS ON ANTHROPOMORPHIC HANDS

| Methods | Type | Hand | Camera | Protocol | Similar/Novel Objects | Success Rate | Time |
|---|---|---|---|---|---|---|---|
| Wei *et al.* [23] | Learning | DLR-HIT | Ensenso N35 | Lift | 10/10 | 74.7% | 0.14s |
| Lu *et al.* [25] | Learning | Allegro | Kinect2 | Lift | 1/7 | 75.0% | 5-10s |
| Ficuciello *et al.* [38] | Learning | Schunk SVH | Asus Xtion PRO LIVE | Lift | 0/3 | 73.0% | N/A |
| Hang *et al.* [6] | Optimization | Allegro | N/A | Shake | 6 | **90.0%** | 1-2s |
| Graspit! [33] | Optimization | Schunk SVH | N/A | Lift & Shake | 16 | 71.9% | 65.3s |
| Ours | Hybrid | Schunk SVH | RealSense D435i | Lift & Shake | 0/16 | 79.7% | **0.033s** |

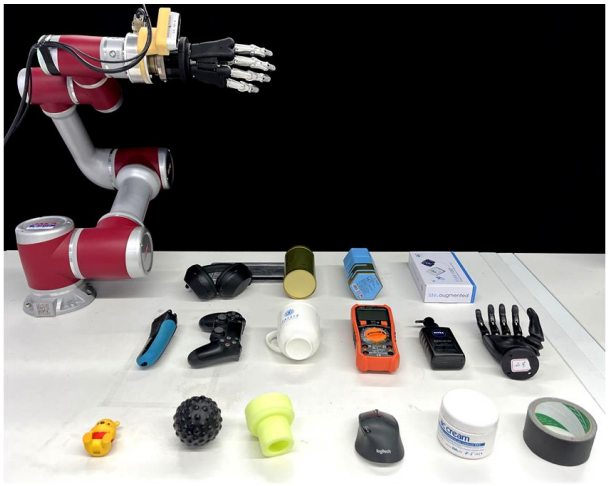

Fig. 8. Real-world experiment setup. The objects from top to bottom are headphone, cylindrical can, hexagonal prism, cardboard box, electric razor, gamepad, mug, multimeter, facial cleanser, prosthetic hand, cartoon model, massage ball, 3D printed object, mouse, cream box, electrical tape.

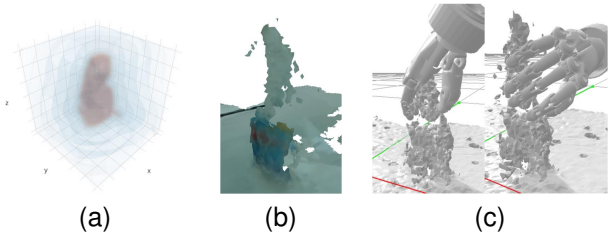


Fig. 9. Grasp generation process of the Windex bottle. (a)SDF volume. (b)Reconstructed mesh. (c)First two grasps generated by our method.

We use sixteen novel objects with various shapes and sizes as shown in Fig. 8. The first collision-free grasp from our approach and the GraspIt are executed. For GraspIt, the captured point cloud is registered with the original object mesh to find the transformation to retrieve grasps from the database. Then the grasps are filtered similarly to Algorithm 1. Each object is grasped eight times with random placements, and 128 grasps are executed. New SDF is integrated from depth images for every new pose of the target object. A successful grasp is lifting the object 10cm and shaking it 10cm horizontally six times at a speed of 10cm/s without dropping it to the table, similar to the benchmark in [41]. Successful grasps synthesized by our method are depicted in Fig.10.

We present the comparison of our method with other optimization-based and learning-based methods for anthropomorphic grasp planning in Table III. We use a larger object set, and a stricter protocol that includes lifting and shaking to introduce external disturbance. Our method outperforms the learning-based methods in both grasp success rate (79.7%) and time efficiency (0.033s), attributing to the lightweight wrist sampling network and force-closure grasp synthesis. Compared with the optimization-based methods, our method is much faster, although [6] has a higher success rate (90%). However, [6] uses six simple precise models of bottles and jars, and requires feedback control based on tactile sensors. We observe that the main reasons for unsuccessful grasps are: positional error of the fingertips from noisy sensory inputs, contact slippage caused by asynchronous finger closure, and insufficient grasping force due to open-loop control.

Fig. 10. Successful grasps generated by our approach on 16 novel objects.

## VI. CONCLUSIONS

This letter presents the FFCGS that produces force-closure grasps for anthropomorphic hands with efficiency. Feasible 6D wrist poses are predicted by a wrist pose sampler to reduce the dimension of search space, followed by a differentiable force-closure grasp synthesis method that find force-closure

grasps efficiently. The simulation experiments show that our approach can propose stable and diverse dexterous grasps in 33ms. Real-world experiments show that our approach can achieve a higher success rate of 79.7% compared to baseline methods. The proposed approach has great potential usage in robotic hand-arm systems and upper-limb prostheses. Future work will focus on solving the limitations of the current work. For example, a reactive grasp controller based on sensorized fingertips to avoid premature contact with the object, which would benefit the success rate and grasp stability.

## REFERENCES


[1] V. Prasad, R. Stock-Homburg, and J. Peters, "Human-Robot Handshaking: A Review," *Int J of Soc Robotics*, vol. 14, no. 1, pp. 277–293, Jan. 2022.

[2] B.-Y. Sun, X. Gong, J. Liang, W.-B. Chen, Z.-L. Xie, C. Liu, and C.-H. Xiong, "Design Principle of a Dual-Actuated Robotic Hand With Anthropomorphic Self-Adaptive Grasping and Dexterous Manipulation Abilities," *IEEE Trans. Robot.*, vol. 38, no. 4, pp. 2322–2340, 2022.

[3] L. Wang, Y. Xiang, and D. Fox, "Manipulation Trajectory Optimization with Online Grasp Synthesis and Selection," in *Robot. Sci. Syst. XVI*, Jul. 2020.

[4] H. Dai, A. Majumdar, and R. Tedrake, "Synthesis and Optimization of Force Closure Grasps via Sequential Semidefinite Programming," in *Robotics Research*, 2018, vol. 2, pp. 285–305.

[5] T. Liu, Z. Liu, Z. Jiao, Y. Zhu, and S.-C. Zhu, "Synthesizing Diverse and Physically Stable Grasps With Arbitrary Hand Structures Using Differentiable Force Closure Estimator," *IEEE Robot. Autom. Lett.*, vol. 7, no. 1, pp. 470–477, 2022.

[6] K. Hang, M. Li, J. A. Stork, Y. Bekiroglu, F. T. Pokorny, A. Billard, and D. Kragic, "Hierarchical Fingertip Space: A Unified Framework for Grasp Planning and In-Hand Grasp Adaptation," *IEEE Trans. Robot.*, vol. 32, no. 4, pp. 960–972, Aug. 2016.

[7] M. Kiatos, S. Malassiotis, and I. Sarantopoulos, "A geometric approach for grasping unknown objects with multifingered hands," *IEEE Trans. Robot.*, vol. 37, no. 3, pp. 735–746, 2020.

[8] L. Shao, F. Ferreira, M. Jorda, V. Nambiar, J. Luo, E. Solowjow, J. A. Ojea, O. Khatib, and J. Bohg, "UniGrasp: Learning a Unified Model to Grasp With Multifingered Robotic Hands," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2286–2293, Apr. 2020.

[9] S. Yu, D.-H. Zhai, Y. Xia, H. Wu, and J. Liao, "SE-ResUNet: A Novel Robotic Grasp Detection Method," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 5238–5245, Apr. 2022.

[10] S. Jain and B. Argall, "Grasp detection for assistive robotic manipulation," in *2016 IEEE Int. Conf. Robot. Autom. ICRA*, 2016, pp. 2015–2021.

[11] X. Zhu and J. Wang, "Synthesis of force-closure grasps on 3-D objects based on the Q distance," *IEEE Trans. Robot. Autom.*, vol. 19, no. 4, pp. 669–679, 2003.

[12] M. Breyer, J. J. Chung, L. Ott, R. Siegwart, and J. Nieto, "Volumetric Grasping Network: Real-time 6 DOF Grasp Detection in Clutter," in *Proc. 2020 Conf. Robot Learn.*, Oct. 2021, pp. 1602–1611.

[13] M. A. Roa and R. Suárez, "Grasp quality measures: Review and performance," *Auton Robot*, vol. 38, no. 1, pp. 65–88, Jan. 2015.

[14] E. Rimon and J. Burdick, "On force and form closure for multiple finger grasps," in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 2, Apr. 1996, pp. 1795–1800 vol.2.

[15] M. Liu, Z. Pan, K. Xu, K. Ganguly, and D. Manocha, "Deep Differentiable Grasp Planner for High-DOF Grippers," in *Robot. Sci. Syst. XVI*, Jul. 2020.

[16] J. D. Schulman, K. Goldberg, and P. Abbeel, "Grasping and Fixturing as Submodular Coverage Problems," in *Robotics Research*, 2017, vol. 100, pp. 571–583.

[17] Y. Fan, X. Zhu, and M. Tomizuka, "Optimization Model for Planning Precision Grasps with Multi-Fingered Hands," in *2019 IEEERSJ Int. Conf. Intell. Robots Syst. IROS*, Nov. 2019, pp. 1548–1554.

[18] T. Osa, "Multimodal trajectory optimization for motion planning," *The International Journal of Robotics Research*, vol. 39, no. 8, pp. 983–1001, Jul. 2020.

[19] A. Mousavian, C. Eppner, and D. Fox, "6-dof graspnet: Variational grasp generation for object manipulation," in *Proc. IEEECVF Int. Conf. Comput. Vis.*, 2019, pp. 2901–2910.

[20] G. Du, K. Wang, S. Lian, and K. Zhao, "Vision-based Robotic Grasping From Object Localization, Object Pose Estimation to Grasp Estimation for Parallel Grippers: A Review," *Artif Intell Rev*, vol. 54, no. 3, pp. 1677–1734, Mar. 2021.

[21] J. Varley, C. DeChant, A. Richardson, J. Ruales, and P. Allen, "Shape completion enabled robotic grasping," in *2017 IEEERSJ Int. Conf. Intell. Robots Syst. IROS*, Sep. 2017, pp. 2442–2447.

[22] J. Lundell, E. Corona, T. Nguyen Le, F. Verdoja, P. Weinzaepfel, G. Rogez, F. Moreno-Noguer, and V. Kyrki, "Multi-FinGAN: Generative Coarse-To-Fine Sampling of Multi-Finger Grasps," in *2021 IEEE Int. Conf. Robot. Autom. ICRA*, 2021, pp. 4495–4501.

[23] W. Wei, D. Li, P. Wang, Y. Li, W. Li, Y. Luo, and J. Zhong, "DVGG: Deep Variational Grasp Generation for Dextrous Manipulation," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 1659–1666, Apr. 2022.

[24] F. Cini, V. Ortenzi, P. Corke, and M. Controzzi, "On the choice of grasp type and location when handing over an object," *Sci. Robot.*, vol. 4, no. 27, p. eaau9757, Feb. 2019.

[25] Q. Lu, M. Van der Merwe, B. Sundaralingam, and T. Hermans, "Multifingered Grasp Planning via Inference in Deep Neural Networks: Outperforming Sampling by Learning Differentiable Models," *IEEE Robot. Autom. Mag.*, vol. 27, no. 2, pp. 55–65, Jun. 2020.

[26] J. Kang, S. Lee, M. Jang, and S. Lee, "Gradient Flow Evolution for 3D Fusion From a Single Depth Sensor," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 2211–2225, Apr. 2022.

[27] P. Mittal, Y.-C. Cheng, M. Singh, and S. Tulsiani, "Autosdf: Shape priors for 3d completion, reconstruction and generation," in *Proc. IEEECVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 306–315.

[28] C. Sommer, L. Sang, D. Schubert, and D. Cremers, "Gradient-SDF: A Semi-Implicit Surface Representation for 3D Reconstruction," in *Proc. IEEECVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 6280–6289.

[29] H. Fu, C. Li, X. Liu, J. Gao, A. Celikyilmaz, and L. Carin, "Cyclical Annealing Schedule: A Simple Approach to Mitigating KL Vanishing," in *Proc. 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Vol. 1 Long Short Pap.*, 2019, pp. 240–250.

[30] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar, "Benchmarking in Manipulation Research: Using the Yale-CMU-Berkeley Object and Model Set," *IEEE Robot. Autom. Mag.*, vol. 22, no. 3, pp. 36–52, Sep. 2015.

[31] D. Morrison, P. Corke, and J. Leitner, "EGAD! An Evolved Grasping Analysis Dataset for Diversity and Reproducibility in Robotic Manipulation," *IEEE Robot. Autom. Lett.*, vol. 5, no. 3, pp. 4368–4375, Jul. 2020.

[32] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," *ArXiv Prepr. ArXiv180109847*, 2018.

[33] A. Miller and P. Allen, "Graspit! A versatile simulator for robotic grasping," *IEEE Robot. Autom. Mag.*, vol. 11, no. 4, pp. 110–122, 2004.

[34] R. D. Howe, I. Kao, and M. R. Cutkosky, "The sliding of robot fingers under combined torsion and shear loading," in *Proc. 1988 IEEE Int. Conf. Robot. Autom.*, 1988, pp. 103–105.

[35] N. Pollard, "Synthesizing grasps from generalized prototypes," in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 3, Apr. 1996, pp. 2124–2130 vol.3.

[36] Y. Fan, H.-C. Lin, T. Tang, and M. Tomizuka, "Grasp Planning for Customized Grippers by Iterative Surface Fitting," in *2018 IEEE 14th Int. Conf. Autom. Sci. Eng. CASE*, Aug. 2018, pp. 28–34.

[37] K. Koyama, M. Shimojo, A. Ming, and M. Ishikawa, "Integrated control of a multiple-degree-of-freedom hand and arm using a reactive architecture based on high-speed proximity sensing," *Int. J. Robot. Res.*, vol. 38, no. 14, pp. 1717–1750, Dec. 2019.

[38] F. Ficuciello, A. Migliozzi, G. Laudante, P. Falco, and B. Siciliano, "Vision-based grasp learning of an anthropomorphic hand-arm system in a synergy-based control framework," *Sci. Robot.*, vol. 4, no. 26, p. eaao4900, Jan. 2019.

[39] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U2-Net: Going deeper with nested U-structure for salient object detection," *Pattern Recognition*, vol. 106, p. 107404, Oct. 2020.

[40] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon, "KinectFusion: Real-time dense surface mapping and tracking," in *2011 10th IEEE Int. Symp. Mix. Augment. Real.*, 2011, pp. 127–136.

[41] Y. Bekiroglu, N. Marturi, M. A. Roa, K. J. M. Adjigble, T. Pardi, C. Grimm, R. Balasubramanian, K. Hang, and R. Stolkin, "Benchmarking Protocol for Grasp Planning Algorithms," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 315–322, Apr. 2020.


**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.**